Local Weighting–Based Diagnostics for Bayesian Poststratification

Ryan Giordano, Alice Cima, Erin Hartman, Jared Murray, Avi Feller

Berkeley BSTARS September 2025

Are US non-voters becoming more Republican?

Blue Rose research says yes:

“Politically disengaged voters have become much more Republican, And because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate.”

(Blue Rose Research 2024) (On Ezra Klein show, major professional pollsters)

***On Data and Democracy* says no:**

“Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available.”

(Bonica et al. 2025) (Berkeley professor co–author, major professional researchers)

Are US non-voters becoming more Republican?

Blue Rose research says yes:

"Politically disengaged voters have become much more Republican, And because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate."

(Blue Rose Research 2024) (On Ezra Klein show, major professional pollsters)

***On Data and Democracy* says no:**

"Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available."

(Bonica et al. 2025) (Berkeley professor co–author, major professional researchers)

- The problem is very hard (it's difficult to accurately poll non–voters)
- Different data sources
- **Very different statistical methods:** ⋆
  - Blue Rose uses Bayesian hierarchical modeling (MrP)
  - The CES uses calibration weighting (CW)

Are US non-voters becoming more Republican?

Blue Rose research says yes:

"Politically disengaged voters have become much more Republican, And because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate."

(Blue Rose Research 2024) (On Ezra Klein show, major professional pollsters)

***On Data and Democracy* says no:**

"Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available."

(Bonica et al. 2025) (Berkeley professor co–author, major professional researchers)

- The problem is very hard (it's difficult to accurately poll non–voters)
- Different data sources
- **Very different statistical methods:** ⋆
  - Blue Rose uses Bayesian hierarchical modeling (MrP)
  - The CES uses calibration weighting (CW)

Our contribution

We define "MrP local equivalent weights" (MrPlew) that:

- Are easily computable from MCMC draws and standard software, and
- Provide MrP versions of key diagnostics that motivate calibration weighting.

⇒ **MrPlew provide apples-to-apples comparisons between MrP and calibration weighting.**

Outline

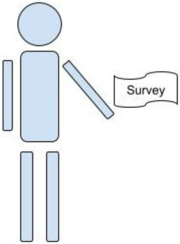
- Introduce the statisical problem and two methods (CW and MrP)
- Describe covariate balance, one of the classical CW diagnostics
- Define MrPlew weights and connect them to covariate balance
- Example of real-world results
- Future directions

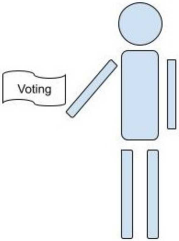
The basic problem

We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to “do you support policy such–and–such”)

We want the average response in a target population, in which we observe only covariates.





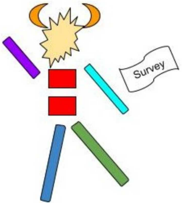
Observe $(\mathbf{x}_i, y_i)$ for $i = 1, \ldots, S$ | Observe $\mathbf{x}_j$ for $j = 1, \ldots, T$

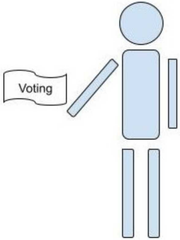
The basic problem

We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to “do you support policy such–and–such”)

We want the average response in a target population, in which we observe only covariates.





Observe $(\mathbf{x}_i, y_i)$ for $i = 1, \ldots, S$ | Observe $\mathbf{x}_j$ for $j = 1, \ldots, T$

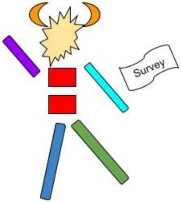
The problem is that the populations are very different.

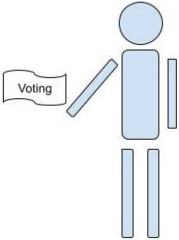
The basic problem

We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to "do you support policy such–and–such")

We want the average response in a target population, in which we observe only covariates.





Observe $(\mathbf{x}_i, y_i)$ for $i = 1, \ldots, S$ | Observe $\mathbf{x}_j$ for $j = 1, \ldots, T$

The problem is that the populations are very different.

Our survey results may be biased.

How can we use the covariates to say something about the target responses?

Two approaches

We want $\mu := \frac{1}{T} \sum_{j=1}^{T} y_j$, but don't observe target population y_j.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Two approaches

We want $\mu := \frac{1}{T}\sum_{j=1}^{T} y_j$, but don't observe target population y_j.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

▶ Choose "calibration weights" w_i (e.g. raking weights)

Bayesian hierarchical modeling (MrP)

▶ Choose model $\mathcal{P}(y|x,\theta)$ and prior $\mathcal{P}(\theta)$ (e.g. Hierarchical logitstic regression)

Two approaches

We want $\mu := \frac{1}{T}\sum_{j=1}^{T} y_j$, but don't observe target population y_j.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

- ▶ Choose "calibration weights" w_i (e.g. raking weights)
- ▶ Take $\hat{\mu}_{\text{CAL}} = \frac{1}{S}\sum_{i=1}^{S} w_i y_i$

Bayesian hierarchical modeling (MrP)

- ▶ Choose model $\mathcal{P}(y|x, \theta)$ and prior $\mathcal{P}(\theta)$ (e.g. Hierarchical logitstic regression)
- ▶ Take $\hat{y}_t = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}[y|\mathbf{x}_j]$ and $\hat{\mu}_{\text{MRP}} = \frac{1}{T}\sum_{j=1}^{T} \hat{y}_j$

Two approaches

We want $\mu := \frac{1}{T}\sum_{j=1}^{T} y_j$, but don't observe target population y_j.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

- Choose "calibration weights" w_i (e.g. raking weights)
- Take $\hat{\mu}_{\text{CAL}} = \frac{1}{S}\sum_{i=1}^{S} w_i y_i$
- Dependence on y_i is obvious (w_s typically chosen using only $\mathbf{x}$)

Bayesian hierarchical modeling (MrP)

- Choose model $\mathcal{P}(y|x,\theta)$ and prior $\mathcal{P}(\theta)$ (e.g. Hierarchical logitstic regression)
- Take $\hat{y}_t = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}\left[y|\mathbf{x}_j\right]$ and $\hat{\mu}_{\text{MRP}} = \frac{1}{T}\sum_{j=1}^{T} \hat{y}_j$
- Dependence on y_s very complicated (Typically via MCMC draws from $\mathcal{P}(\theta|\text{Survey data})$)

Two approaches

We want $\mu := \frac{1}{T}\sum_{j=1}^{T} y_j$, but don't observe target population y_j.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

- ▶ Choose "calibration weights" w_i (e.g. raking weights)
- ▶ Take $\hat{\mu}_{\text{CAL}} = \frac{1}{S}\sum_{i=1}^{S} w_i y_i$
- ▶ Dependence on y_i is obvious (w_s typically chosen using only $\mathbf{x}$)
- ▶ Weights give interpretable diagnostics:
  - Frequentist variability
  - Partial pooling
  - Regressor balance

Bayesian hierarchical modeling (MrP)

- ▶ Choose model $\mathcal{P}(y|x,\theta)$ and prior $\mathcal{P}(\theta)$ (e.g. Hierarchical logitstic regression)
- ▶ Take $\hat{y}_t = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}\left[y|\mathbf{x}_j\right]$ and $\hat{\mu}_{\text{MRP}} = \frac{1}{T}\sum_{j=1}^{T}\hat{y}_j$
- ▶ Dependence on y_s very complicated (Typically via MCMC draws from $\mathcal{P}(\theta|\text{Survey data})$)
- ▶ **Black box**

Two approaches

We want $\mu := \frac{1}{T}\sum_{j=1}^{T} y_j$, but don't observe target population y_j.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

▶ Choose "calibration weights" w_i (e.g. raking weights)

▶ Take $\hat{\mu}_{\text{CAL}} = \frac{1}{S}\sum_{i=1}^{S} w_i y_i$

▶ Dependence on y_i is obvious (w_s typically chosen using only $\mathbf{x}$)

▶ Weights give interpretable diagnostics:

- Frequentist variability
- Partial pooling
- Regressor balance

Bayesian hierarchical modeling (MrP)

▶ Choose model $\mathcal{P}(y|x,\theta)$ and prior $\mathcal{P}(\theta)$ (e.g. Hierarchical logitstic regression)

▶ Take $\hat{y}_t = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}[y|\mathbf{x}_j]$ and $\hat{\mu}_{\text{MRP}} = \frac{1}{T}\sum_{j=1}^{T}\hat{y}_j$

▶ Dependence on y_s very complicated (Typically via MCMC draws from $\mathcal{P}(\theta|\text{Survey data})$)

▶ **Black box**

$\leftarrow$ (We open this box, providing analogues of all these diagnostics)

What are we weighting for?[1]

We want:

$$\text{Target average response} = \frac{1}{T}\sum_{j=1}^{T} y_j \approx \frac{1}{S}\sum_{i=1}^{S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j.

[1] Pun attributable to Solon, Haider, and Wooldridge (2015)

What are we weighting for?[1]

We want:

$$\text{Target average response} = \frac{1}{T}\sum_{j=1}^{T} y_j \approx \frac{1}{S}\sum_{i=1}^{S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j. But we can check whether:

$$\frac{1}{T}\sum_{j=1}^{T} \mathbf{x}_j = \frac{1}{S}\sum_{i=1}^{S} w_i \mathbf{x}_i$$

Such weights satisfy "covariate balance" for $\mathbf{x}$.

You can check covariate balance for any calibration weighting estimator, and any function $f(\mathbf{x})$.

[1] Pun attributable to Solon, Haider, and Wooldridge (2015)

What are we weighting for?[1]

We want:

$$\text{Target average response} = \frac{1}{T}\sum_{j=1}^{T} y_j \approx \frac{1}{S}\sum_{i=1}^{S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j. But we can check whether:

$$\frac{1}{T}\sum_{j=1}^{T} \mathbf{x}_j = \frac{1}{S}\sum_{i=1}^{S} w_i \mathbf{x}_i$$

Such weights satisfy "covariate balance" for **x**.

You can check covariate balance for any calibration weighting estimator, and any function $f(\mathbf{x})$.

Even more, covariate balance is the criterion for a popular class of calibration weight estimators:

Raking calibration weights

"Raking" selects weights that

- Are as "close as possible" to some reference weights
- Under the constraint that they balance some selected regressors.

[1] Pun attributable to Solon, Haider, and Wooldridge (2015)

Generalized covariate balance checks

We want to balance $f(\mathbf{x})$ because we think $\mathbb{E}[y|\mathbf{x}]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Generalized covariate balance checks

We want to balance $f(\mathbf{x})$ because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Balance–informed sensitivity check (BISC) (informal)

Pick a small δ, and define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the change this is supposed to induce in the target population.

Covariate balance checks whether our estimators produce the same change.

Generalized covariate balance checks

We want to balance $f(\mathbf{x})$ because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Balance–informed sensitivity check (BISC) (formal)

Pick a small δ, and define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the expected change this perturbation produces in the target distribution:

$$\mathbb{E}\left[\mu(\tilde{y}) - \mu(y)|\mathbf{x}\right] = \frac{1}{T}\sum_{j=1}^{T}\left(\mathbb{E}\left[\tilde{y}|\mathbf{x}_p\right] - \mathbb{E}\left[y|\mathbf{x}_p\right]\right) = \delta\frac{1}{T}\sum_{j=1}^{T} f(\mathbf{x}_j)$$

Then, check whether your estimator $\hat{\mu}(\cdot)$ produces the same change for observed $\tilde{y}, y$:

$$\underbrace{\hat{\mu}(\tilde{y}) - \hat{\mu}(y)}_{\text{Replace weighted averages with changes in an estimator}} \overset{\text{check}}{\approx} \delta\frac{1}{T}\sum_{j=1}^{T} f(\mathbf{x}_j).$$

Generalized covariate balance checks

We want to balance $f(\mathbf{x})$ because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Balance–informed sensitivity check (BISC) (formal)

Pick a small δ, and define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the expected change this perturbation produces in the target distribution:

$$\mathbb{E}\left[\mu(\tilde{y}) - \mu(y)|\mathbf{x}\right] = \frac{1}{T}\sum_{j=1}^{T}\left(\mathbb{E}\left[\tilde{y}|\mathbf{x}_p\right] - \mathbb{E}\left[y|\mathbf{x}_p\right]\right) = \delta\frac{1}{T}\sum_{j=1}^{T} f(\mathbf{x}_j)$$

Then, check whether your estimator $\hat{\mu}(\cdot)$ produces the same change for observed $\tilde{y}, y$:

$$\underbrace{\hat{\mu}(\tilde{y}) - \hat{\mu}(y)}_{\text{Replace weighted averages with changes in an estimator}} \overset{\text{check}}{\approx} \delta\frac{1}{T}\sum_{j=1}^{T} f(\mathbf{x}_j).$$

When $\hat{\mu}(\cdot) = \hat{\mu}_{\mathrm{CAL}}(\cdot)$, BISC recovers the standard covariate balance check.

When $\hat{\mu}(\cdot) = \hat{\mu}_{\mathrm{MRP}}(\cdot)$ and δ is small, BISC recovers our proposal.

Generalized covariate balance for MrP

Step one: Construct $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.

Generalized covariate balance for MrP

Step one: Construct $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.

Problem: Our y is binary! (We're motivated by hierarchical linear regression.)

Generalized covariate balance for MrP

Step one: Construct $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.

Problem: Our y is binary! (We're motivated by hierarchical linear regression.)

Two possibilities:

- Allow $\tilde{y}$ to take values other than $\{0, 1\}$ and set $\tilde{y} = y + \delta f(\mathbf{x})$, or
- Use an estimate of $\mathbb{E}\left[y|\mathbf{x}\right]$ to draw new binary $\tilde{y}$.

Our approach:

- Use $\tilde{y} = y + \delta f(\mathbf{x})$ to identify problematic "imbalanced" $f(\mathbf{x})$
- Sanity check by generating binary $\tilde{y}$ using $f(\mathbf{x})$ (which is fast and easy)

Generalized covariate balance for MrP

Step one: Construct $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.

Generalized covariate balance for MrP

Step one: Construct $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Step two: Evaluate $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$.

Generalized covariate balance for MrP

Step one: Construct $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Step two: Evaluate $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$.

Problem: $\hat{\mu}_{\text{MRP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- Output is noisy, and $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$ may be small.

Generalized covariate balance for MrP

Step one: Construct $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Step two: Evaluate $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$.

Problem: $\hat{\mu}_{\text{MRP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- Output is noisy, and $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$ may be small.

MrP Local Equivalent Weights (MrPlew)

Form the approximation

$$\hat{\mu}_{\text{MRP}}(\tilde{y}) = \sum_{i=1}^{S} w_i^{\text{MRP}}(\tilde{y}_i - y_i) + \text{Residual} \quad \text{where} \quad w_i^{\text{MRP}} := \frac{d}{dy_i}\hat{\mu}_{\text{MRP}}(y).$$

Generalized covariate balance for MrP

Step one: Construct $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Step two: Evaluate $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$.

Problem: $\hat{\mu}_{\text{MRP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- Output is noisy, and $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$ may be small.

MrP Local Equivalent Weights (MrPlew)

Form the approximation

$$\hat{\mu}_{\text{MRP}}(\tilde{y}) = \sum_{i=1}^{S} w_i^{\text{MRP}} (\tilde{y}_i - y_i) + \text{Residual} \quad \text{where} \quad w_i^{\text{MRP}} := \frac{d}{dy_i} \hat{\mu}_{\text{MRP}}(y).$$

The weights are given by weighted averages of posterior covariances (Giordano, Broderick, and Jordan 2018).

They can be easily computed with standard software[2] **without re–running MCMC**.

[2]We use `brms` (Bürkner 2017).

Generalized covariate balance for MrP

Step one: Construct $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Step two: Evaluate $\hat{\mu}_{\mathrm{MRP}}(\tilde{y}) - \hat{\mu}(y)$.

Problem: $\hat{\mu}_{\mathrm{MRP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- Output is noisy, and $\hat{\mu}_{\mathrm{MRP}}(\tilde{y}) - \hat{\mu}(y)$ may be small.

MrP Local Equivalent Weights (MrPlew)

Form the approximation

$$\hat{\mu}_{\mathrm{MRP}}(\tilde{y}) = \sum_{i=1}^{S} w_i^{\mathrm{MRP}} (\tilde{y}_i - y_i) + \text{Residual} \quad \text{where} \quad w_i^{\mathrm{MRP}} := \frac{d}{dy_i} \hat{\mu}_{\mathrm{MRP}}(y).$$

We state conditions under which, as $\delta \to 0$, and $N \to \infty$,
the residual is of lower order than the MrPlew term, *uniformly in* $f(\cdot)$.

Based on prior work on uniform and finite–sample error bounds for Bernstein–von Mises theorem–like results (Giordano and Broderick 2024).

(See also Kasprzak, Giordano, and Broderick (2025)!)

Generalized covariate balance for MrP

Step one: Construct $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Step two: Evaluate $\hat{\mu}_{\mathrm{MRP}}(\tilde{y}) - \hat{\mu}(y)$.

Problem: $\hat{\mu}_{\mathrm{MRP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- Output is noisy, and $\hat{\mu}_{\mathrm{MRP}}(\tilde{y}) - \hat{\mu}(y)$ may be small.

MrP Local Equivalent Weights (MrPlew)

Form the approximation

$$\hat{\mu}_{\mathrm{MRP}}(\tilde{y}) = \sum_{i=1}^{S} w_i^{\mathrm{MRP}}(\tilde{y}_i - y_i) + \text{Residual} \quad \text{where} \quad w_i^{\mathrm{MRP}} := \frac{d}{dy_i}\hat{\mu}_{\mathrm{MRP}}(y).$$

If MrP were linear (e.g. if you use OLS instead of hierarchical logistic regression), then

- The residual is zero,
- $\hat{\mu}_{\mathrm{MRP}}(y) = \sum_{i=1}^{S} w_i^{\mathrm{MRP}} y_i$, and so
- $\hat{\mu}_{\mathrm{MRP}}(\tilde{y})$ is a calibration weighting estimator, and w_i^{MRP} are its weights. (Cite Gelman)

In general, MrP is truly nonlinear. The residual is only small when $\tilde{y} \approx y$ (i.e., when $\delta \ll 1$).

Covariate balance

Theorem

- Let $\tilde{y} = y + \delta f(\mathbf{x})$,
- $\hat{\mu}_{\mathrm{MRP}}$ be a hierachical logistic regression posterior expectation, and
- $\mathcal{F}$ be a Donsker class of uniformly bounded functions on $\mathbf{x}$.

Then, with probability approaching one, as $N \to \infty$,

$$\sup_{f \in \mathcal{F}} \left(\hat{\mu}_{\mathrm{MRP}}(\tilde{y}) - \left(\hat{\mu}_{\mathrm{MRP}}(y) + \sum_{i=1}^{S} w_s^{\mathrm{MRP}} \delta f(\mathbf{x}_s) \right) \right) = O(\delta^2) \quad \text{as } \delta \to 0$$

The supremum over $\mathcal{F}$ is the primary technical contribution! It means we are justified in searching over regressors to find imbalance.

Draws on our prior work on uniform and finite–sample error bounds for Bernstein–von Mises theorem–like results (Giordano and Broderick 2024; Kasprzak, Giordano, and Broderick 2025).

Real Data

Analysis of changing names after marriage (based on Alexander (2019))

- **Target population:** ACS survey of US population 2017–2022 (Ruggles et al. 2024))
- **Survey population:** Marital Name Change Survey (Cohen 2019)
- **Respose:** Did the female partner keep their name after marriage?
- For regressors, use bins of age, education, state, and decade married.

Survey observations:	$S = 4,364$
Target observations (rows):	$T = 4,085,282$

Uncorrected survey mean:	$\frac{1}{S}\sum_{i=1}^{S} y_n = 0.462$
Raking:	$\hat{\mu}_{\text{CAL}} = 0.263$
MrP:	$\hat{\mu}_{\text{MRP}} = 0.288 \quad (\text{Post. sd} = 0.0169)$

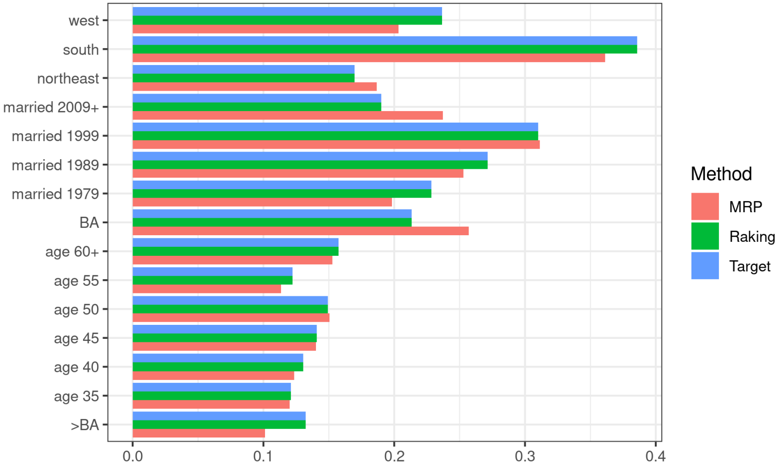


Figure 1: Imbalance plot for primary effects

Figure

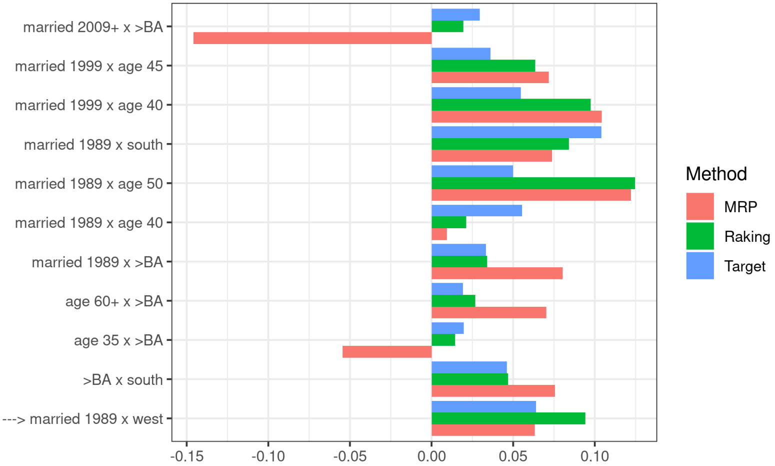


Figure 2: Imbalance plot for select interaction effects

Figure

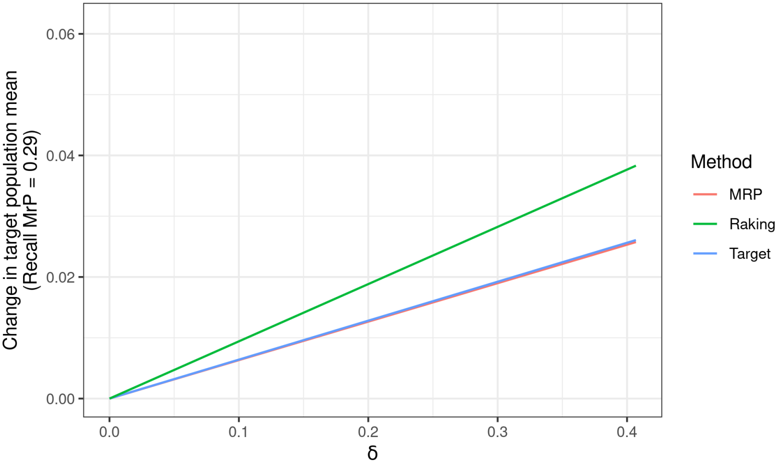


Figure 3: Continuous predictions Alexander

Figure

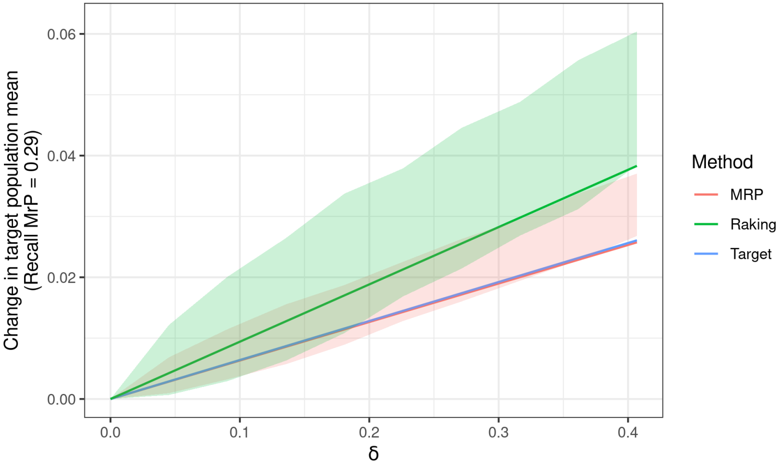


Figure 4: Continuous predictions Alexander

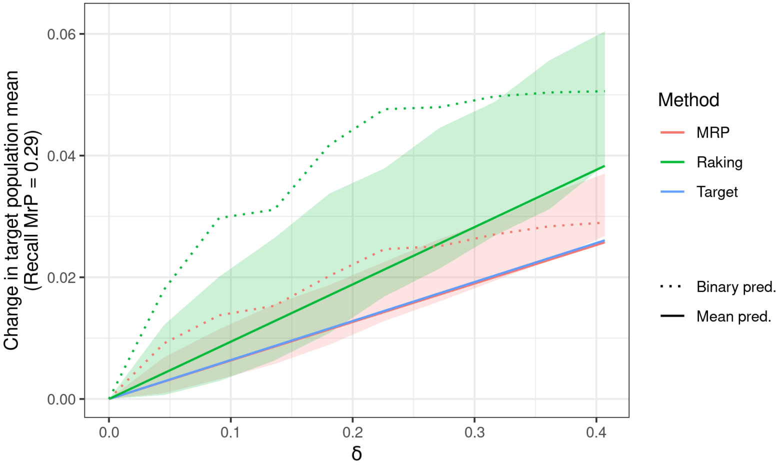


Figure 5: Continuous predictions Alexander

Figure

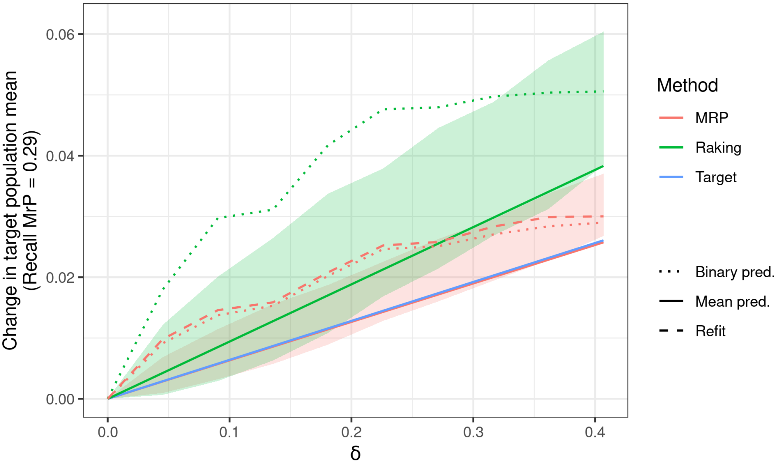


Figure 6: Continuous predictions Alexander

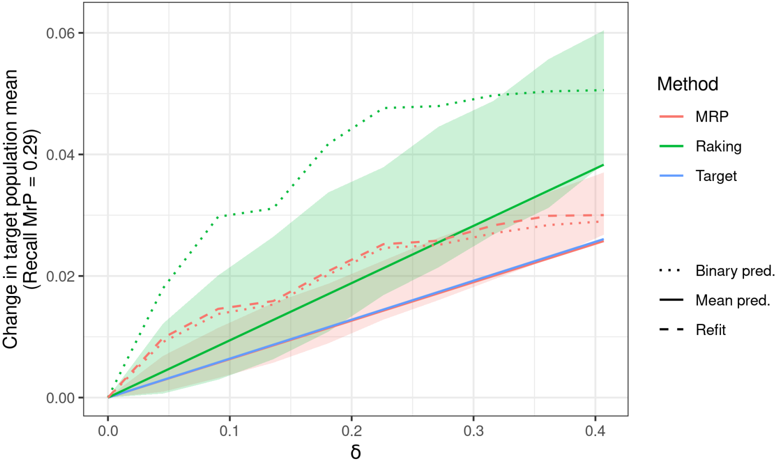


Figure 6: Continuous predictions Alexander

Running ten MCMC refits: 28 hours Computing approximate weights: 27 seconds

Future work and generalizations

- Instance of a very general class of local consistency checks that generalize classical regression checks (work with Sequoia)
- Versions for GLMMs (work with Vladimir)
- Going beyond classical Bayesian sensitivity (work with Lucas)

References

- Alexander, M. (2019). *Analyzing name changes after marriage using a non-representative survey*. URL: https://www.monicaalexander.com/posts/2019-08-07-mrp/.
- Blue Rose Research (2024). *2024 Election Retrospective Presentation*. https://data.blueroseresearch.org/2024retro-download. Accessed on 2024-10-26.
- Bonica, A. et al. (Apr. 2025). *Did Non-Voters Really Flip Republican in 2024? The Evidence Says No.* https://data4democracy.substack.com/p/did-non-voters-really-flip-republican.
- Bürkner, Paul-Christian (2017). "brms: An R Package for Bayesian Multilevel Models Using Stan". In: *Journal of Statistical Software* 80.1, pp. 1–28. DOI: 10.18637/jss.v080.i01.
- Cohen, P. (Apr. 2019). *Marital Name Change Survey*. DOI: 10.17605/OSF.IO/UZQDN. URL: osf.io/uzqdn.
- Giordano, R. and T. Broderick (2024). *The Bayesian Infinitesimal Jackknife for Variance*. arXiv: 2305.06466 [stat.ME]. URL: https://arxiv.org/abs/2305.06466.
- Giordano, R., T. Broderick, and M. I. Jordan (2018). "Covariances, robustness and variational bayes". In: *Journal of machine learning research* 19.51.
- Kasprzak, M., R. Giordano, and T. Broderick (2025). *How good is your Laplace*